

一种新的扩展模糊概念数据模型

秦艳友¹ 赵法信^{1,2}

(通化师范学院 通化 134002)¹ (东北大学信息科学与工程学院 沈阳 110004)²

摘要 为了对现实世界中存在的不精确、不确定的信息进行建模,学者们提出了各种不同的扩展关系数据模型。基于可能性理论,本文对概念数据模型 IFO 的不同层次在模糊信息环境下进行了扩展,提出了一种新的模糊 IFO 数据模型 FIFO,该模型能够有效地表达和处理概念数据建模层次中的模糊数据。文章重点讨论了 FIFO 中对象和关系(主要是 ISA 关系)的模糊性,并给出了 FIFO 模型的各个层次所对应的图形表达。

关键词 模糊集,可能性理论,数据建模,IFO 模型,FIFO 模型

A New Extended Fuzzy Conceptual Data Model

QIN Yan-You¹ ZHAO Fa-Xin^{1,2}

(Tonghua Normal University, Tonghua 134002)¹

(School of Information Science and Engineering, Northeastern University, Shenyang 110004)²

Abstract In order to model the real world with imprecise and uncertain information, various extended relational data models are proposed. Based on possibility theory, different levels of fuzziness are introduced into conceptual data model IFO and then a new fuzzy IFO model, denoted as FIFO, is proposed. The model can effectively represent and process fuzzy data at conceptual data modeling level. The paper concentrates on the fuzzification of objects and relationships in the FIFO, specially on that of ISA relationships, and the corresponding graphical representations of the different levels in FIFO are given.

Keywords Fuzzy set, Possibility theory, Data modeling, IFO model, FIFO model

1 引言

数据库已经历了从层次和网状数据库到关系数据的发展。但是随着 CAD/CAM、多媒体以及知识系统等用于非事务性处理的计算机技术的出现,关系数据库的局限性也日益显现出来。为此,研究者们提出了一些非关系数据模型(如实体-关系数据模型,面向对象数据模型和逻辑数据模型)作为已有数据库或潜在数据库模型的建模工具,这些数据库模型统称为后关系数据模型。可以说,现实世界的需求加速了数据库技术的发展,将附加的语义引入数据模型已经成为数据库研究的一个主要目标。

在现实世界中,信息经常是模糊的或不精确的。然而,包括关系数据库和后关系数据模型在内的经典数据库只能建模和处理完备的信息。为了满足对不精确和不确定信息进行建模的需求,各种不同种类的不精确信息已经被引入到关系数据库中^[1,2],而最近的研究已经侧重于在面向对象数据库中,对不精确和不确定信息进行建模^[3,4]。相对于不精确和不确定关系数据库和面向对象数据库方面的广泛研究,对不精确和不确定概念数据建模方面的研究则很少。

文[5]根据模糊集理论,在 ER 数据模型中的属性层、实例/模式层以及模式层三个层次中引入了模糊信息。Chen 和 Kerre 则于 1998 年^[6]将该模型模糊扩展至 EER 数据模型。除了 ER 和 EER 数据模型外,Abiteboul 和 Hull 在文[7]中提出了另一种概念数据模型,即 IFO 数据模型。该模型将语义数据库建模的基本原理融入到基于图形的表达框架中。与其

它数据库模型一样,IFO 数据模型也不能直接对模糊信息进行建模。为了实现模糊关系数据库的概念设计,文[8]对 IFO 模型进行了扩展,将不同层次的模糊信息引入到原子对象类型和关联中,并给出了将一个模糊 IFO 模式映射到一个模糊关系模式的机制。然而,文[8]并未对两种构造方法和 ISA 关系进行模糊扩展。

本文基于 Abiteboul 和 Hull^[7]所提出的框架,将模糊信息引入到 IFO 数据模型的各个不同的层次,并给出了相应的图形表达,从而将其扩展至模糊 IFO 数据模型 FIFO,进而着重对 FIFO 数据模型中的对象和关系(主要是 ISA 关系)进行了讨论。

2 基础知识

2.1 IFO 模型

IFO 数据模型^[7]是一种用数学定义的概念数据模型,它将语义数据库建模的基本原理融入到基于图形的表达框架中。形式上讲,一个 IFO 模式就是一个由各种类型的顶点和边组成的用于表示原子对象、构造对象、关联以及 ISA 关系的有向图。一个基本的 IFO 模式就由这几种元素构成。

2.1.1 对象

各种不同对象结构的表达方法被称为类型,它是构成任意 IFO 模式的基础,并与模式框图中的节点相对应。在 IFO 中存在三种原子类型。这三种类型还可以进一步构造以产生更为复杂的类型。

IFO 模型中存在的三种原子类型如下:

- (1)基本类型,此类型对应于对象的预定义数据类型。
 (2)抽象类型,此类型对应于现实世界中没有基本结构的对象。
 (3)自由类型,此类型对应于通过 ISA 关系获得的实体。
 这三种原子类型所对应的图形表达如下所示。

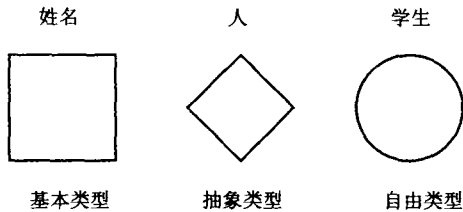


图1 IFO模型的三种原子类型

非原子对象可以通过使用以下两种方法对原子类型进行构造获得:

- (1)组合:用于描述给定结构中有限对象的集合。
 (2)聚合:用于形成有序的实例 n 元组,它与一个类型相关联。

而且,这两种构造方法还可以以任意顺序进行递归以形成更加复杂的类型。图2给出了这两种构造方法所对应的图形表达。

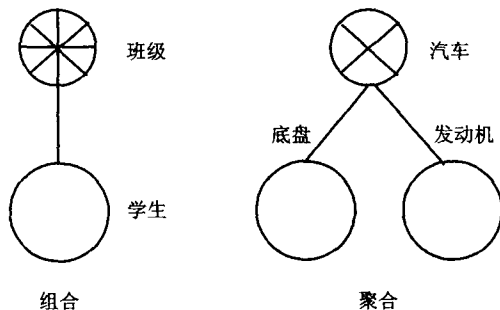


图2 IFO对象类型的两种构造方法

2.1.2 关联

IFO模型的另一个主要结构化组件是关联,它用于表示函数关系。关联自然地提供了各种类型的聚合方式以及它们之间的关联函数。从图3所示的关联中可以看出,函数“驾驶”将一个“汽车”映射到该“汽车”的“驾驶员”,函数“驾驶时间”将每一个“驾驶员”和他们对该车的驾驶“时间”对应起来。在这里,“驾驶员”、“引擎”和“底盘”是自由类型,“汽车”是用聚合方法构建的构造对象类型,“时间”则是基本类型。

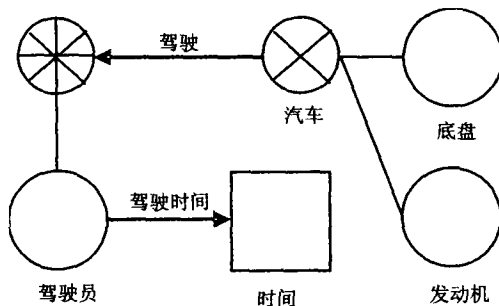


图3 一个关联实例

2.1.3 ISA关系

IFO模型的最后一个结构化组件是ISA关系的表达,它主要有两种表达方法:

(1)概化,用一个粗箭头表示,主要适用于表达将已存在的不同类型合并为一个虚拟类型的情形。

(2)特化,用一个双线箭头表示,可用来为一个给定类型中的成员定义可能的角色。这种情况主要适用于属性继承。

概化和特化的图形表达如图4所示。

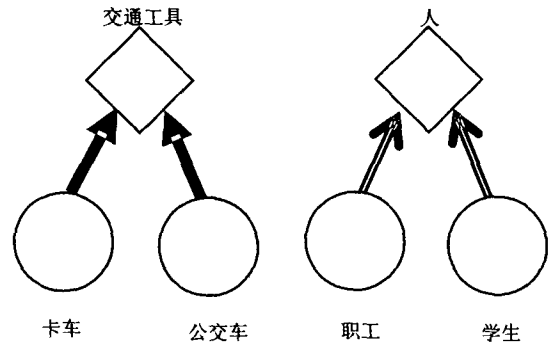


图4 概化与特化实例

利用上述的这些基本组件就可以合成IFO模式。但是,传统的IFO模型不能对广泛存在于现实世界中的不精确和不确定信息进行建模。

2.2 模糊集和可能性分布

Zadeh最早用模糊集表示模糊数据。设 U 为论域, U 上的模糊集 F 表示 U 上的模糊值,模糊集 F 上的隶属函数定义为 $\mu_F: U \rightarrow [0,1]$, 其中 $\mu_F(u) (\forall u \in U)$ 表示 u 对模糊集 F 的隶属度。因此模糊集 F 可表示为:

$$F = \{\mu(u_1)/u_1, \mu(u_2)/u_2, \dots, \mu(u_n)/u_n\}.$$

在此方法中,当 $\mu_F(u)$ 被解释成变量 X 取 U 中的值 u 的可能性时,模糊值可以用可能性分布 π_X 表示^[9]。设 π_X 和 F 分别为模糊值的可能性分布和模糊集表示,很明显有 $\pi_X = F$ 成立。

为了用模糊集表示多值属性中的不完备值,本文将模糊集分为析取模糊集和合取模糊集两种。其中析取模糊集是指只能从多个值中选择唯一值,合取模糊集则可以选择一个或多个值。

3 FIFO数据模型

为了对模糊信息进行建模,本文将语义模型IFO模糊扩展为FIFO数据模型。下面,我们将讨论如何在属性层、实例/模式层以及模式层对IFO进行模糊扩展。

3.1 模糊基本类型

在FIFO模型中,属性层的模糊性可以用模糊基本类型来表示。对于客观世界中的模糊基本类型对象,它的实例中的相应属性必定有一个模糊值。在这里我们将模糊基本类型分为析取模糊基本类型和合取模糊基本类型两种,析取模糊基本类型意味着只能从多个选项中选择唯一值,而合取模糊基本类型则意味着可以从多个选项中选择多个值。此外,在数据模型中,基本类型可以是模糊地隶属于一个对象,并用一个隶属度值表示其隶属于该对象的程度。在FIFO模型中,我们可以通过在基本类型框图中放置隶属度值来表示这样的模糊基本类型。

析取模糊基本类型、合取模糊基本类型以及具有隶属度值的模糊基本类型所对应的图形表达如图5所示。

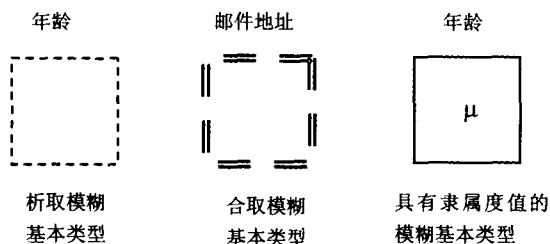


图5 三种模糊基本 FIFO 类型

3.2 模糊抽象类型和模糊自由类型

抽象类型和自由类型的模糊性体现在实例/模式层和模式层这两个层次的模糊性上。实例/模式层的模糊性与特定对象的实例相关,这意味着对象实例分别模糊地隶属于相应的对象。例如,张三可能是个博士生。而模式层的模糊性则意味着对象可以模糊地隶属于相应的数据模型,且有一个值表示其隶属于该数据模型的程度。我们可以在 FIFO 模型中的抽象类型框图和自由类型框图中放置隶属度值。例如,设 A 是一个抽象类型, μ 是它在模型中的隶属度,则将“ μ ”($0 < \mu \leq 1$)放置在菱形框内,当 $\mu=1$ 时,可以将“1”略去不写。模糊抽象类型和模糊自由类型在实例/模式层和模式层的图形表达如图 6 所示。

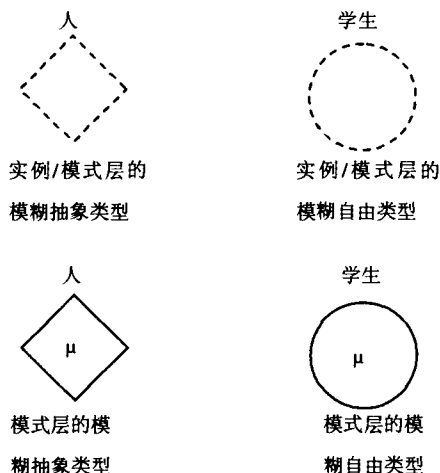


图6 四种模糊抽象和模糊自由 FIFO 类型

3.3 模糊构造方法

首先,我们看一下聚合构造方法,这个构造器将表示对象各部分的子类型连接成表示整个对象的类型,从而形成高一层级的对象。这个子类型可以是上面提到的完备的或模糊的原子类型,也可以是通过聚合或组合而得到的构造类型。当参与聚合构造的子类型中存在带有隶属度的模糊类型时,其相应的聚合结果也是一个带有隶属度的模糊类型。该隶属度值是所有参与聚合的子类型隶属度的乘积。

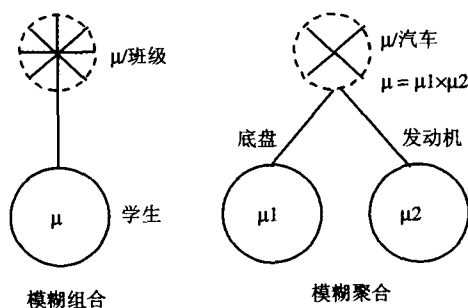


图7 FIFO 中对象类型的两种模糊构造方法

与聚合构造器相似,当参与组合构造的任一子类型是带有隶属度的模糊类型时,其相应的组合结果也是一个带有隶属度的组合值。该模糊组合的隶属度即为参与组合的子类型的隶属度。

带有隶属度的模糊组合和模糊聚合所对应的图形表达如图 7 所示。

3.4 模糊关联

在模糊环境下,抽象与抽象类型、抽象与自由类型以及自由与自由类型之间关联可以是模糊的。对这种模糊性共有两种解释,一种解释是对象之间的函数关系是确定的,但函数关系的具体实例是模糊的。图 8(A)给出了这种模糊关联的图形表达。另一种解释是对象之间的函数关系是模糊的,但这种模糊关联必须带有一个隶属度值。该模糊关联所对应的图形表达如图 8(B)所示。

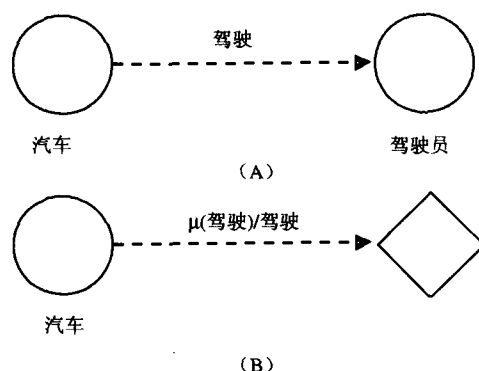


图8 两种模糊关联实例

3.5 模糊 ISA 关系

FIFO 模型中的 ISA 关系与子类/超类的概念相关。设 $E, S_1, S_2, \dots$ 和 S_n 是 FIFO 模型中的非基本对象类型。当且仅当在 $E, S_1, S_2, \dots$ 和 S_n 的实例/模式层存在模糊性且表达式 $(\forall e)(\forall S)(S \in \{S_1, S_2, \dots, S_n\} \wedge \mu_S(e) \leq \mu_E(e))$ 其中 e 为论域中的对象实例成立时,我们才可以说 $S_1, S_2, \dots$ 和 S_n 是 E 的模糊子类, E 是 $S_1, S_2, \dots$ 和 S_n 的模糊超类。

图 9 所示为 FIFO 模型中模糊概化和模糊特化的图形表达:

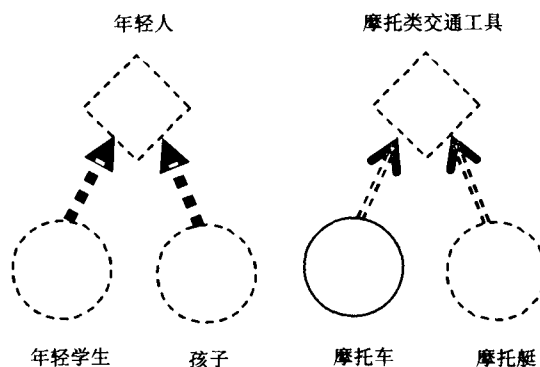


图9 模糊概化与模糊特化实例

3.6 一个 FIFO 模型实例

下面是一个车间的实例。一个车间由设备和工人组成,设备可以特化为新设备和旧设备,在这里新设备和旧设备是实例/模式层具有模糊性的模糊类型。车间中工人的年龄可

(下转第 222 页)

3.2 实验及结果

本文的实验在奔 4 1.7G CPU 和 512MM 存储器上运行。初始训练样本数 n' 为 30, 域值 a_i 取 0.15, 常数 K 取 3.1, 升级鲁棒尺度参数 σ 的调节常数 $\beta=2$, 增量升级权值比例 $\alpha=0.95$ 。为了验证本文算法的有效性, 本文从运行时间和重建质量两方面比较标准 PCA, Torre et al^[1] 的鲁棒 PCA 方法和本文的鲁棒增量 PCA 算法。由于本文算法和 Torre 方法能自动获得基样本数, 本文初始鲁棒参数估计自动得到的初始基图像数为 18, 为了公平起见, 标准 PCA 主成分数也设为 18, 由三种算法得到的部分基图像和重建效果如图 2 所示。

由图 2 可知, 标准 PCA 重建结果受人活动引起的遮挡和光照明显变化, 局部强污染等情况影响较大, Torre 方法虽然能处理遮挡和光照明显变化的情况, 但对局部强污染效果也不理想。本文的方法由于引入样本级局外点排除过程, 取得了较好的重建效果。从定量上来分析重建质量, 三种方法对整个训练集的均值平方重建误差 (MSRE 定义为 $(1/n) \sum_{i=1}^n \|d_i - \mu - Bc_i\|_2^2$) 分别为: 2.2129, 0.05626 和 0.04925。从算法运行时间来看, Torre 的鲁棒算法运行需要 3h10m, 而本文的鲁棒增量子空间算法仅需要 18m, 大大快于 Torre 的鲁棒方法。本文的鲁棒增量算法最费时的部分是在采用迭代的鲁棒初始参数学习部分, 费时约 16m, 增量部分能处理图像约 2fps。由图 2 可知, 本文算法能捕获光照变化, 得到鲁棒的子空间模型, 又由于能排除样本级局外点, 如图 2(e) 所示, 可以应用在自适应背景建模中, 进一步来检测运动对象, 这是下一步需要研究的内容。

结论 本文提出了一种新的鲁棒子空间建模方法, 该方

法主要利用基于双平方函数的鲁棒估计, 梯度下降学习规则和 M-估计器来学习和估计初始参数, 自动从样本级和信号级两个层次来排除局外点的影响, 然后采用一个考虑到样本级和信号级局外点的鲁棒增量方式来学习更新参数, 获得可靠的子空间模型。试验表明, 本文的鲁棒子空间算法, 能够有效处理各种不同类型的噪音, 在背景建模时可以获得反映场景亮度变化的子空间模型, 能有效解决传统背景建模对亮度明显变化, 遮挡, 噪音污染等敏感问题, 同时具有较快的在线学习速度。下一步研究方向有: 进行适量修改, 将本文的方法扩展到其它子空间学习技术中; 在自适应背景建模中应用本文方法, 进一步用来检测运动对象。

参考文献

- 1 De La Torre, Black M J. A Framework for Robust Subspace Learning. *International Journal of Computer Vision*, 2003, 54 (1): 117~142
- 2 Coates T F, Edwards G J, Taylor C J. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(6): 681~685
- 3 Diamantaras K I. *Principal Component Neural Networks (Theory and Applications)*. John Wiley & Sons, 1996
- 4 Black M J, Jepson A D. Eigentracking: Robust matching and tracking of objects using view-based representation. *International Journal of Computer Vision*, 1998, 26(1): 63~84
- 5 Xu L, Yuille A. Robust principal component analysis by self-organizing rules based on statistical physics approach. *IEEE Transactions on Neural Networks*, 1995, 6(1): 131~143
- 6 Skocaj D. *Robust Subspace Approaches to Visual Learning and Recognition*. University of Ljubljana, Doctoral Dissertation, 2003
- 7 Li Y. On incremental and robust subspace learning. *Pattern Recognition*, 2004, (37): 1509~1518
- 8 DuMouchel W, Brien F O. Integrating a Robust Option into a Multiple Regression Computing Environment. *American Statistical Association*, Alexandria, VA, 1989. 297~301

(上接第 198 页)

以是模糊的。另外, 从可能性值 0.2 可以看出, 工人很少在旧设备上进行操作。利用在本节中介绍的有关图形符号, 就可以将上述状态描述用 FIFO 数据模型表示出来, 如图 10 所示。

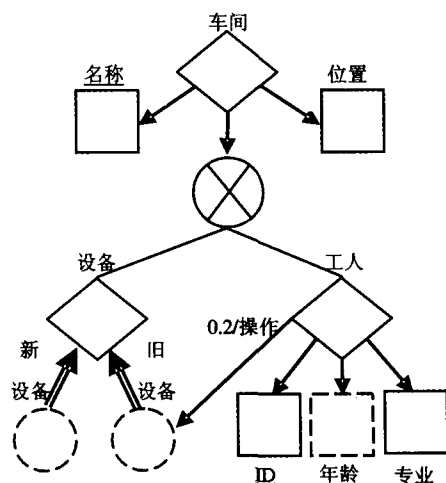


图 10 一个 FIFO 数据模型实例

结束语 计算机技术在非传统领域的应用需要对概念数据进行建模。然而, 已有的概念数据模型在实现日益增长的现实应用时具有一定的局限性。对现实世界中的模糊信息进行建模就是这些应用之一。实际上, 模糊性是现实世界的一

个特点。而在当前, 针对模糊概念数据模型进行建模的研究则很少。

本文将模糊信息引入了 IFO 数据模型的各个不同层次。并给出了 FIFO 所对应的图形表达, 根据本文所提出的概念和方法, 我们能够有效地表达和处理概念数据建模层次中的模糊数据。

参考文献

- 1 Rundensteiner E A, Hawkes L W, Bandler W. On Nearness Measures in Fuzzy Relational Data Models [J]. *International Journal of Approximate Reasoning*, 1989, 3: 267~298
- 2 Sheno S, Melton A. Proximity Relations in the Fuzzy Relational Databases [J]. *Fuzzy Sets and Systems*, 1989, 31(3): 285~296
- 3 Gysegheem N V, Caluwe R D. Imprecision and Uncertainty in UFO Database Model [J]. *Journal of the American Society for Information Science*, 1998, 49(3): 236~252
- 4 Bordogna G, Pasi G, Lucarella D. A Fuzzy Object-Oriented Data Model for Managing Vague and Uncertain Information [J]. *International Journal of Intelligent Systems*, 1999, 14: 623~651
- 5 Zvieli A, Chen P P. Entity-Relationship Modeling and Fuzzy Databases [C]. In: *Proceedings of the 1986 IEEE International Conference on Data Engineering*, 1986. 320~327
- 6 Chen G Q, Kerre E E. Extending ER/EER Concepts towards Fuzzy Conceptual Data Modeling [C]. In: *Proceedings of the 1998 IEEE International Conference on Fuzzy Systems*, 1998, 2: 1320~1325
- 7 Abiteboul S, Hull R. IFO: A Formal Semantic Database Model [J]. *ACM Transactions on Database Systems*, 1987, 12(4): 525~565
- 8 Ma Z M, Zhang W J, Ma W Y. Conceptual Design of Relational Databases with Imprecise and Uncertain Information [C]. In: *Proceedings of the Fourth Asian Fuzzy Systems Symposium*, 2000. 869~874
- 9 Zadeh L A. Fuzzy Sets as a Basis for a Theory of Possibility [J]. *Fuzzy Sets and Systems*, 1978, 1(1): 3~28